

T-Test, ANOVA and Chi-Square Test

A student-friendly lecture on selecting, applying, interpreting, and reporting common inferential statistical tests

Author: Muhammad Hassan
Website: ResearchMethod.net

Lecture Overview

Why these tests matter

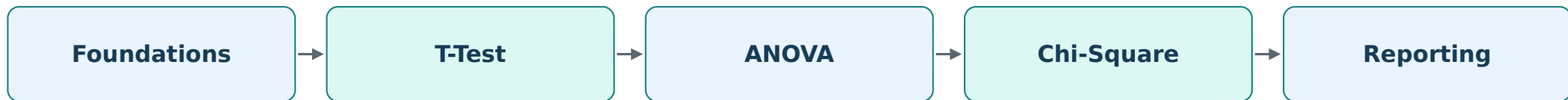
T-tests, ANOVA, and chi-square tests help researchers decide whether observed differences or relationships are likely to reflect real patterns rather than random sampling variation.

What the lecture covers

Concepts, assumptions, test selection, step-by-step analysis, interpretation, reporting, examples, ethics, and classroom practice.

How to use the slides

Each test family is explained with practical examples and simple rules that students can apply in research projects.



Core question: Which statistical test matches the research question, variables, design, and assumptions?

The lecture moves from “what test?” to “how do I interpret it?”

Learning Objectives

Distinguish between t-tests, ANOVA, and chi-square tests.

Select an appropriate statistical test based on variables and research design.

Explain null and alternative hypotheses in simple research language.

Identify key assumptions for each test family.

Interpret p-values, test statistics, confidence intervals, and effect sizes.

Report results clearly using academic writing conventions.

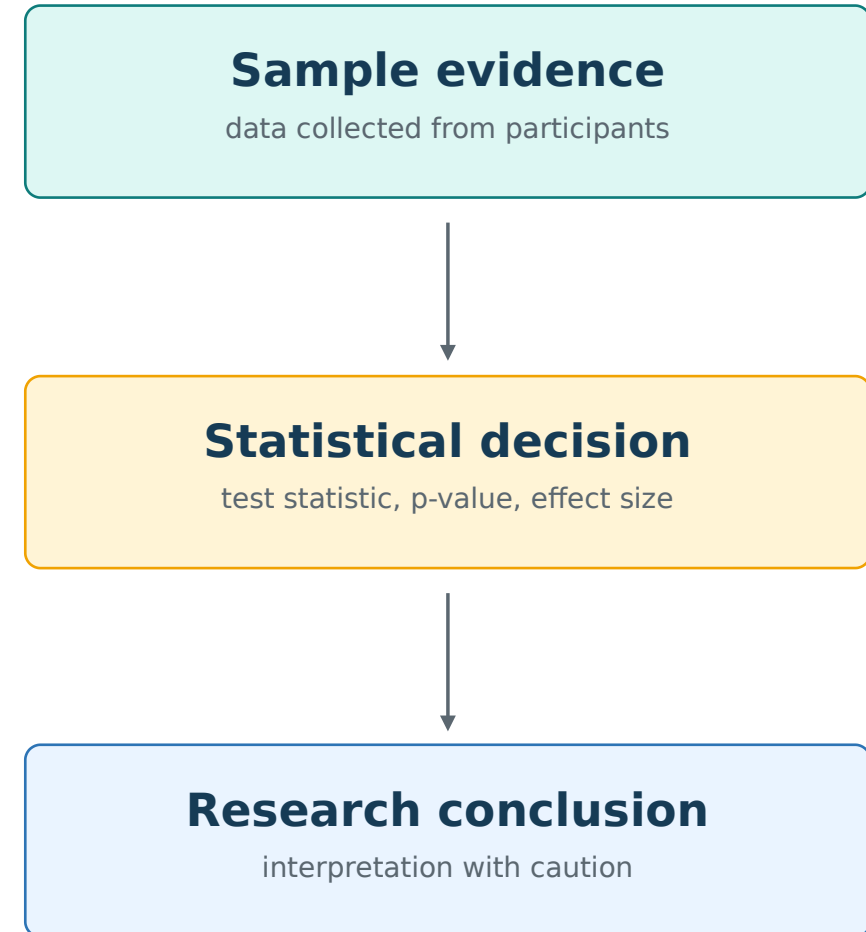
Recognize ethical and practical issues in statistical analysis.

By the end, students should be able to answer:

“Which test should I use, why, and what do the results mean?”

Introduction: Why Inferential Tests Are Used

Researchers usually study a sample, then make careful claims about a wider population. Inferential tests estimate whether a result is large enough to be unlikely under the null hypothesis. T-tests and ANOVA focus on mean differences for quantitative outcomes. Chi-square tests focus on categorical counts and patterns of association. Correct test selection supports credible conclusions and prevents misleading claims.



Key Definitions

Term	Student-friendly meaning
Null hypothesis (H_0)	The default statement: no difference, no association, or no effect.
Alternative hypothesis (H_1)	The research expectation: a difference, association, or effect exists.
p-value	How surprising the sample result would be if H_0 were true.
α level	The chosen cut-off for statistical significance, commonly .05.
Effect size	How large or meaningful the difference or association is.
Assumption	A condition that should be reasonably met for a test to be valid.

Tip: “Statistically significant” means the data are unlikely under H_0 ; it does not automatically mean “large,” “important,” or “causal.”

Background and Context

Student's t-test

Developed from early work by William Sealy Gosset, published under the name "Student," to analyze small-sample mean differences.

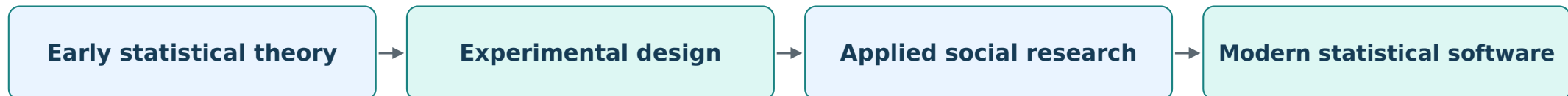
ANOVA

Associated with Ronald A. Fisher's work on experimental design and comparing multiple group means efficiently.

Chi-square

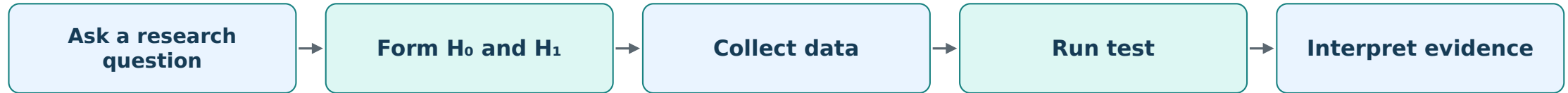
Based on Karl Pearson's work on comparing observed counts with expected counts under a statistical model.

Modern research uses these tests in software such as SPSS, R, Stata, Jamovi, JASP, Excel, Python, and SAS.



The tests are widely used because they provide a clear structure for testing claims from sample data.

Main Concepts: The Logic of Statistical Testing



Evidence

A test statistic summarizes the difference between what was observed and what would be expected under the null hypothesis.

Uncertainty

The p-value helps judge whether the result is compatible with random sampling variation.

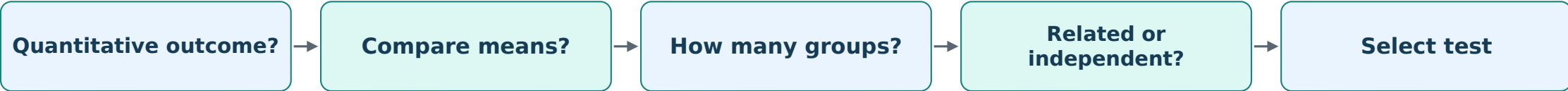
Meaning

Effect size and context help judge whether the result matters in practical or theoretical terms.

Good analysis = correct test + checked assumptions + cautious interpretation + transparent reporting

Choosing the Right Test

Start with the outcome variable and the research question.



Research situation	Typical test
One quantitative sample vs known value	One-sample t-test
Two independent groups, quantitative outcome	Independent-samples t-test
Same participants measured twice	Paired-samples t-test
Three or more independent group means	One-way ANOVA
Two categorical variables	Chi-square test of independence
Observed counts vs expected distribution	Chi-square goodness-of-fit

T-Test: Purpose and Logic

A t-test compares a mean or two means when the outcome variable is quantitative.

It asks whether an observed mean difference is large relative to the variation in the data.

Common versions: one-sample, independent-samples, and paired-samples t-tests.

Typical output includes t value, degrees of freedom, p-value, confidence interval, and effect size.

$$\frac{\text{t statistic}}{\frac{\text{difference between means}}{\text{standard error}}}$$

If the difference is large compared with sampling error, the t value becomes larger.

Example: Do students who receive tutoring have a different average exam score than students who do not?

One-Sample T-Test

Purpose

Tests whether the mean of one sample differs from a known or hypothesized value.

Example question

Is the average stress score of university students different from the standard score of 50?

Variables

One quantitative outcome measured in one sample; one comparison value.

Element	Illustration
H ₀	Mean stress score = 50
H ₁	Mean stress score ≠ 50
Data	Stress scores from sampled students
Interpretation	Is the sample mean meaningfully different from 50?

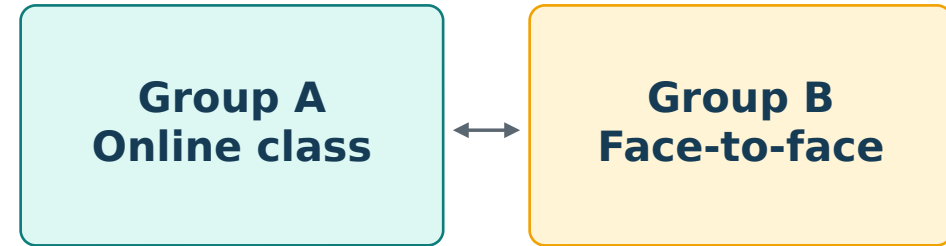
Independent-Samples T-Test

Used when comparing the means of two separate groups.

The groups should be independent: a participant belongs to only one group.

Often used in experiments and cross-sectional comparison studies.

The outcome must be quantitative; the grouping variable has two categories.



Compare mean achievement scores

Output to report

t(df), p-value, mean difference, confidence interval, and effect size such as Cohen's d.

Paired-Samples T-Test

Used when two scores are linked for each participant or case.

Common in pre-test/post-test designs, repeated measurements, and matched pairs.

The test analyzes the mean of the difference scores. The key question is whether the average change is different from zero.



Example: Did anxiety scores decrease after a mindfulness program?

Important distinction

The data are not independent because each post-test score is connected to a specific pre-test score.

T-Test Assumptions and Diagnostics

Assumption	Why it matters	How to check
Quantitative outcome	Means must be meaningful	Confirm measurement scale
Independence	Observations should not influence each other	Review sampling/design
Approximately normal differences or scores	Supports t distribution inference	Histogram, Q-Q plot, sample size
Homogeneity of variance	For independent t-test only	Levene’s test; compare SDs
No extreme outliers	Outliers can distort means	Boxplots; data audit

When assumptions are seriously violated, consider transformations, robust methods, non-parametric alternatives, or clearer limitations.

ANOVA: Purpose and Logic

ANOVA stands for Analysis of Variance.

It compares means across three or more groups or conditions.

It tests whether at least one group mean differs from another.

ANOVA avoids running many t-tests, which increases Type I error risk.

Common output includes F value, degrees of freedom, p-value, and effect size.

ANOVA compares:
variation between groups

variation within groups

Large between-group variation relative to within-group variation leads to a larger F value.

One-Way ANOVA

Purpose

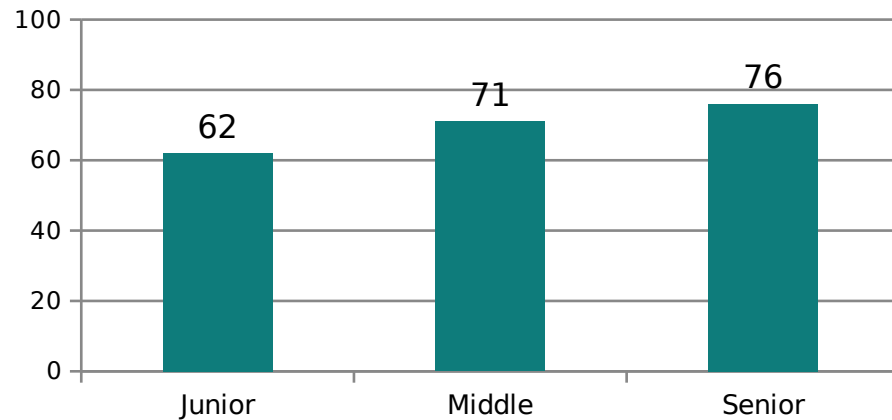
Compares the means of three or more independent groups on one quantitative outcome.

Factor

One categorical independent variable with three or more levels.

Example

Mean job satisfaction across junior, middle, and senior employees.



Illustrative group means

H_0 : All group means are equal.

H_1 : At least one group mean is different.

If significant, use post-hoc comparisons when appropriate.

Two-Way and Factorial ANOVA

Factorial ANOVA examines two or more categorical independent variables.
It can test main effects: the separate effect of each factor.
It can test interaction effects: whether the effect of one factor depends on another factor.
Useful when research questions involve combinations of conditions.

Factor B: Teaching method	Factor A: Low motivation	Factor A: High motivation
Online	Mean score A	Mean score B
Face-to-face	Mean score C	Mean score D

Example research question

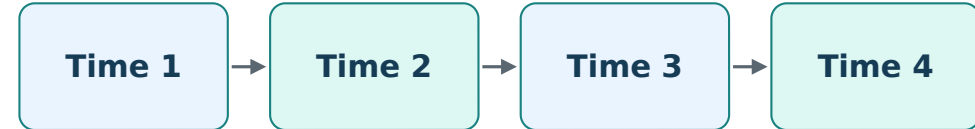
Does teaching method affect achievement, and does that effect differ for students with low versus high motivation?

Repeated-Measures ANOVA

Used when the same participants are measured across three or more time points or conditions. Extends the paired-samples t-test beyond two related measurements.

Common in longitudinal, clinical, psychology, education, and usability studies.

Important assumptions include normality of difference patterns and sphericity.



Example: Compare average depression scores before treatment, after treatment, and at follow-up.

Key idea

The repeated scores are correlated because they come from the same people.

ANOVA Assumptions, Post-Hoc Tests, and Effect Size

Area	What to consider
Assumptions	Independent observations, quantitative outcome, approximate normality, homogeneity of variance
Post-hoc tests	Used after significant ANOVA to identify which groups differ
Planned contrasts	Specific comparisons defined before analysis
Effect size	Eta squared (η^2), partial eta squared, or omega squared indicate magnitude
Interpretation	Report both statistical evidence and substantive meaning

A significant ANOVA tells us that not all means are equal; it does not automatically identify every pairwise difference.

Chi-Square Test: Purpose and Logic

Chi-square tests analyze categorical data using counts or frequencies.

They compare observed counts with expected counts under the null hypothesis.

They do not compare means of quantitative variables. Common versions: goodness-of-fit and test of independence.

Output includes χ^2 value, degrees of freedom, p-value, and an effect size such as Cramér's V.

χ^2 asks:

Are the observed counts close to the expected counts?

Counts must represent cases, not percentages alone.

Chi-Square Goodness-of-Fit Test

Tests whether one categorical variable follows an expected distribution.
Used when categories are compared with theoretical, historical, or expected proportions.
Example: Are students’ preferred learning formats equally distributed across online, hybrid, and face-to-face options?
 H_0 states that observed category counts match the expected distribution.

Category	Observed	Expected
Online	48	40
Hybrid	39	40
Face-to-face	33	40

Compare observed counts with expected counts in the same categories.

Chi-Square Test of Independence

Tests whether two categorical variables are associated.
Data are organized in a contingency table.
 H_0 states that the variables are independent.
A significant result suggests the distribution of one variable differs across categories of the other variable.

Learning format	Satisfied	Not satisfied
Online	60	25
Hybrid	70	20
Face-to-face	80	15

Example question: Is learning format associated with satisfaction level?

Interpretation focus

Look at patterns in the table, not only the p-value.

Chi-Square Assumptions and Effect Sizes

Issue	Guidance
Data type	Use counts/frequencies for categorical variables.
Independence	Each case should contribute to only one cell.
Expected counts	Expected cell counts should generally not be too small.
Sample size	Very large samples can make small differences significant.
Effect size	Use phi for 2×2 tables or Cramér’s V for larger tables.

Practical interpretation should include the table pattern and effect size, especially when the sample is large.

Step-by-Step Process for Applying the Tests

1. State the research question and hypotheses.
2. Identify variable types: quantitative or categorical.
3. Identify the design: one group, two groups, three or more groups, repeated measures, or categories.
4. Choose the statistical test that matches the design.
5. Screen data and check assumptions.
6. Run the test and review output.
7. Interpret p-value, confidence interval, effect size, and context.
8. Report the result transparently with limitations.



Decision rule

The test must fit the data—not the other way around.

Instructor tip

Ask students to justify the selected test before calculating results.

Practical Examples Across Disciplines

Discipline	Research example	Likely test
Social sciences	Mean political efficacy between two groups	Independent t-test
Business	Customer satisfaction across three service channels	One-way ANOVA
Education	Reading scores before and after intervention	Paired t-test
Health sciences	Treatment adherence by gender and age group	Chi-square independence
Psychology	Stress score across four therapy conditions	ANOVA
Information technology	User success rate by device type	Chi-square independence

Examples should always be checked against the exact variables, sample design, and assumptions.

Detailed Research Scenario

Research topic

Teaching format and student achievement in an undergraduate research methods course.

Research problem

The department wants to know whether achievement differs across online, hybrid, and face-to-face sections.

Research question

Do mean final exam scores differ by teaching format?

Element	Scenario detail
Participants	Students enrolled in three course formats
Data collection	Final exam score and teaching format recorded after the course
Variables	Exam score = quantitative outcome; teaching format = three-group categorical factor
Data analysis method	One-way ANOVA; post-hoc tests if overall result is significant
Expected interpretation	Determine whether at least one teaching format differs in mean achievement; report effect size and caution about non-random assignment.

Reporting and Interpretation

Report the test name, test statistic, degrees of freedom, p-value, and effect size.

Use exact p-values when possible, such as $p = .032$, rather than only $p < .05$.

Include descriptive statistics: means and standard deviations, or counts and percentages.

Interpret findings in relation to the research question.

Avoid saying that the null hypothesis is “proven.”

T-test template

Group A scored higher than Group B, $t(df) = x.xx$, $p = .xxx$, $d = x.xx$.

ANOVA template

Mean scores differed by group, $F(df_1, df_2) = x.xx$, $p = .xxx$, $\eta^2 = .xx$.

Chi-square template

The variables were associated, $\chi^2(df, N = x) = x.xx$, $p = .xxx$, $V = .xx$.

Comparison Table: T-Test, ANOVA, and Chi-Square

Feature	T-test	ANOVA	Chi-square
Main purpose	Compare one or two means	Compare three or more means	Analyze categorical counts
Outcome variable	Quantitative	Quantitative	Categorical counts
Grouping variable	None, two groups, or paired time points	Three or more groups/factors	One or two categorical variables
Main statistic	t	F	χ^2
Common effect size	Cohen's d	η^2 / partial η^2	Phi / Cramér's V
Example	Two teaching methods	Four training programs	Gender by voting preference

Advantages and Applications

Accessible

Widely taught and available in nearly all statistical software packages.

Research-friendly

Fit many common research designs in education, health, psychology, business, and social science.

Clear reporting

Produce concise statistics that can be reported in theses, articles, and reports.

Support evidence-based comparison and association testing.
Help researchers move from raw data to defensible conclusions.
Encourage transparent decision-making through hypotheses and assumptions.
Can be combined with confidence intervals and effect sizes for richer interpretation.

Limitations and Challenges

Results can be misleading when assumptions are ignored.

A significant p-value does not prove causation or practical importance.

Small samples may have low statistical power; very large samples can make trivial effects significant.

Outliers, missing data, measurement error, and biased samples can distort conclusions.

Chi-square tests do not show the direction or cause of an association without careful table interpretation.

ANOVA may require post-hoc tests to clarify which groups differ.

Caution

**Statistical tests cannot
repair weak research
design, poor measurement,
or biased sampling.**

Common Mistakes

Choosing a test after looking for the most significant result.

Using a t-test repeatedly instead of ANOVA for three or more groups.

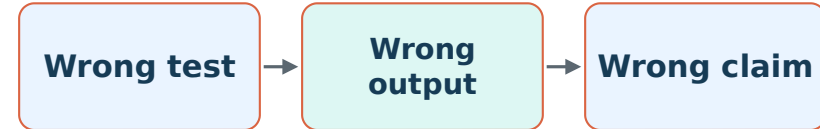
Using chi-square with percentages when the underlying counts are unavailable.

Ignoring independence, normality, homogeneity of variance, or expected count assumptions.

Reporting only p-values without descriptive statistics or effect sizes.

Interpreting association as causation in non-experimental studies.

Not checking data entry errors, outliers, or missing data patterns.



Best quick fix

Write a simple analysis plan before running the test: variables, groups, hypothesis, assumptions, and reporting format.

Best Practices

1. Match the test to the research question and variable types.
2. Describe the sample, groups, and measurement clearly.
3. Check assumptions before interpreting results.
4. Report descriptive statistics with inferential statistics.
5. Use confidence intervals and effect sizes alongside p-values.
6. Visualize data where helpful: boxplots, means plots, or contingency tables.
7. Explain limitations and avoid causal claims unless the design supports them.

Before analysis

Plan the test and assumptions.

During analysis

Check data quality and diagnostics.

After analysis

Interpret cautiously and report transparently.

Ethical Considerations

Protect participant privacy when using quantitative datasets.

Use informed consent and explain how data will be analyzed when appropriate.

Avoid p-hacking, selective reporting, or changing hypotheses after seeing results without disclosure.

Report non-significant findings honestly when they answer the research question.

Do not manipulate categories, remove outliers, or exclude cases only to improve significance.

Share analysis decisions, limitations, and conflicts of interest transparently.

Ethical statistics

**Integrity means
presenting the evidence
fairly, even when results
are unexpected.**

Activity or Classroom Exercise

Activity objective

Practice selecting the correct test and justifying the decision.

Instructions

In groups, read each mini-scenario, identify the variables, and choose t-test, ANOVA, or chi-square.

Expected outcome

Students explain the chosen test using variable type, number of groups, and design.

Mini-scenario	Student task
Compare mean sleep hours between male and female students	Select test and justify
Compare mean satisfaction across four departments	Select test and justify
Check whether device type is associated with purchase decision	Select test and justify
Compare anxiety scores before and after counseling	Select test and justify

Discussion Questions

Why is test selection more than memorizing formulas?

What could happen if a researcher runs many t-tests instead of ANOVA?

How can a result be statistically significant but not practically important?

Why does chi-square require counts rather than only percentages?

How should researchers respond when assumptions are violated?

What ethical problems can occur during statistical analysis?

Why should effect size be reported with p-values?

How can study design limit causal interpretation?

Knowledge Check

#	Question
1	Which test compares the mean of one sample with a known value?
2	Which test is suitable for comparing three independent group means?
3	Which test analyzes the association between two categorical variables?
4	What does a p-value help researchers judge?
5	Name one assumption of the independent-samples t-test.
6	Why are effect sizes important?
7	What does a significant ANOVA result tell us?
8	Why should researchers avoid interpreting correlation or association as causation?

Answer Key, Key Takeaways, and Conclusion

#	Answer
1	One-sample t-test.
2	One-way ANOVA.
3	Chi-square test of independence.
4	How surprising the result is if the null hypothesis is true.
5	Examples: independence, quantitative outcome, approximate normality, equal variances.
6	They show the magnitude or practical meaning of a result.
7	At least one group mean differs; post-hoc tests may identify where.
8	Association does not prove direction, mechanism, or control of confounders.

T-tests compare one or two means.
ANOVA compares multiple means and can test complex designs.
Chi-square tests analyze categorical count data.
Interpretation should combine p-values, effect sizes, assumptions, and research context.
Good statistical practice improves the credibility of research conclusions.

Conclusion: These tests are essential tools for making evidence-based comparisons and association claims, but they must be selected, checked, and reported responsibly.

References

- Agresti, A. (2018). Statistical methods for the social sciences (5th ed.). Pearson.
- American Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.). American Psychological Association.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Routledge.
- Field, A. (2018). Discovering statistics using IBM SPSS Statistics (5th ed.). SAGE Publications.
- Fisher, R. A. (1925). Statistical methods for research workers. Oliver and Boyd.
- Gosset, W. S. (1908). The probable error of a mean. Biometrika, 6(1), 1–25. <https://doi.org/10.2307/2331554>
- Howell, D. C. (2013). Statistical methods for psychology (8th ed.). Wadsworth.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science. Frontiers in Psychology, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302), 157–175. <https://doi.org/10.1080/14786440009463897>

Thank You

Author: Muhammad Hassan
Website: ResearchMethod.net

Visit ResearchMethod.net for research guides, templates, lecture slides, examples, and academic resources.